ONLINE STATISTICS FOR AUSTRALIAN, NEW ZEALAND AND ASIAN ACADEMIC LIBRARIES

Cathie Jilovsky
CAVAL

ABSTRACT
This paper describes the Australian and New Zealand Academic and Research Library Statistics and a pilot project for the benchmarking of library statistics for Asian Academic libraries. CAVAL Collaborative Solutions, an Australian library consortium, is managing the pilot project with sponsorship being provided by the iGroup, Asia.

CAVAL provides a range of services to its members and customers. These include the collection and publication of library statistics and the provision of consultancy services and professional development and training in Statistics and Library Assessment. In particular CAVAL has managed the collection of the Australian Academic and Research Library Statistics for CAUL (Council of Australian University Librarians) statistics since 1992. In 2004 CAVAL explored options for the development of an online statistics website for the CAUL Statistics and a pilot site was developed, through CAVAL’s partnership with the ARL (Association of Research Libraries) and modelled on the well-known ARL online statistics site. Following enthusiastic evaluation by CAUL libraries the pilot site was further developed and the service went live in 2005. Since that time a number of enhancements have been added to the CAUL Statistics Online site.

The objectives of the Asian Statistics pilot project are to provide the same functionality as the CAUL Statistics online website for Asian academic libraries; to improve the collection processes for the individual libraries and to develop a sustainable service for statistical benchmarking.

Phase 1 of the pilot, during which 2 years worth of data is being collected, loaded and made available for benchmarking from the online website, is almost complete. Twenty-two libraries from four countries in the region – Malaysia, Singapore, Thailand and Hong Kong – are participating. Phase 2 will allow for the collection of an additional years data from the participating libraries. The online website and the service will then be evaluated by the participating libraries, the iGroup and CAVAL.

The paper concludes with an outline of some of the challenges and issues which have arisen through the Asian libraries pilot project.

There are many reasons for modern libraries to gather, store, analyse, and provide access to data about their users, operations, personnel, services and collections. Efficient and effective tools are essential in order to make better business and service decisions, to make the library more visible and to increase user satisfaction. These
tools may include applications such as decision support systems, online analytical processing and statistical analysis, forecasting, and data mining. CAVAL has developed an online statistical benchmarking service which provides immediate results to the desktop and gives individuals the capability to manipulate data in a sophisticated fashion.

CAVAL is a public not-for-profit company owned by 12 universities across the Australian states of Victoria, New South Wales and Tasmania. Established in 1978 as a cooperative venture by the Victorian academic libraries, CAVAL provides a variety of services to libraries on a collaborative and commercial basis including shared catalogue systems and services, consultancy services, training and staff development, inter-library-loan and document delivery services and systems, and storage for library materials. These services include the collection and publication of library statistics for several library sectors which is enhanced through the provision of consultancy services and professional development and training in Statistics and Library Assessment.

The Australian and New Zealand Academic and Research Library Statistics have been managed by CAVAL for CAUL (Council of Australian University Librarians) since 1992. More recently CAVAL has been working with a number of Asian Academic libraries on a pilot project for the benchmarking of library statistics in the Asian region with sponsorship being provided by the iGroup, Asia.

**CAUL STATISTICS**

Statistics have been collected annually for the Australian University Library community since 1953, and were originally published in the ‘News Sheet of the University and College Libraries Section, Library Association of Australia’. Data from 1953 to 1983 was published in the ‘Red Book’ of Library Statistics (Long, 1986). New Zealand University library data has been included with the CAUL Statistics since 1974, this is coordinated through CONZUL (Council of New Zealand University Librarians).

Until 1996 the CAUL Statistics data was collected using printed forms and the data was then keyed into a mainframe SPSS database from which the output reports were generated and printed externally. This process was streamlined in 1996 when the use of Excel spreadsheets and email were introduced. Contributing libraries completed an Excel spreadsheet that incorporated the data definitions and a template for the data values to be entered. The CAUL Statistics website was established in 1997 in order to make the collated statistical data easily accessible and publicly available. Since that time the website has expanded to incorporate the data definitions and Excel templates for each year’s statistical collection, along with links to other relevant statistical websites and the CAUL Statistics reports back to 1993 (those prior to 1996 were converted from SPSS to Excel spreadsheet format).
In 2003 the CAUL Statistics Focus Group conducted a survey of CAUL members to find out how the statistics are used, and gain an understanding of requirements for further development of the statistics. There was strong support for a number of enhancements, especially for an online statistics website. The development of the CAUL online statistics website followed a pilot facilitated through CAVAL’s partnership with the ARL (Association of Research Libraries) and modelled on the well-known ARL online statistics site, managed and hosted by the University of Virginia. This was achieved in mid 2004, and the site was then tested and reviewed by CAUL staff responsible for data collection – their response was enthusiastic. Following the pilot CAUL engaged CAVAL to develop and implement the site as an ongoing service.

The site, at http://statistics.caul.edu.au, went live in early 2005. Data back to 1995 has since been loaded into the online database. The online statistics site includes a data entry module which has been used for collection of the data since 2004. The website was further developed and a number of enhancements incorporated into the site in 2006 and 2007. These have included:

- online display of Comments/Notes fields supplied by contributing institutions
- data which is an estimate only is displayed in italics
- maximum number of institutions which can be displayed increased to 15
- facility to select data by Region added – regions available include each Australian state, New Zealand and CAUL
- a ‘Standard Set’, which calculates the ratios and rankings previously provided in the annual printed version of the CAUL Statistics

The CAUL Statistics were published annually in print form in the journal AARL (Australian Academic and Research Libraries) until 2006 when the decision was made to only provide the data electronically. Data from 1995 onwards is available through the CAUL Online Statistics website and data prior to 1995 is available in Excel spreadsheet form from the CAUL statistics website.

The dimensions of the data collection have changed considerably over the 55 years. The number of contributing institutions increased with the creation of new universities and colleges in the 1960’s and 1970’s, and then reduced following the cycle of mergers and amalgamations which took place in the late 1980’s and early 1990’s. The number of data elements reached a peak in 1992, this was reduced over the following decade to make the collection more manageable and keep it relevant.

The following table summarises these changes.

The design and development of the online database raised some complex issues in relation to data relationships and definitions. For example, what degree of change to the name of a data element and/or its associated definition is required before it should be regarded as a different element? It was a relatively easy decision to make the current item ‘Non-serial items: total’ equivalent to the older ‘Total monograph items in library’ and even older ‘Total monograph volumes in li-
library’ data items. However the former ‘Most common number of opening hours during semester’ and ‘Most common number of opening hours during vacation’ are clearly not equivalent to the current ‘Opening Hours’. Similar issues apply to institutional names, although it can be argued that a name change is simply a label, whereas a merger or amalgamation creates a new organisation. In the past some (but not all) multi-campus University libraries supplied data for each campus, whereas nowadays data is only collected for the library as a whole. Each of these relationships needed to be clearly defined when mapping the data into the online database. A decision was made to discontinue the use of column numbers in the online system. Although column numbers have been a useful shortcut reference to data elements, they cause confusion when items are added or removed over time.

<table>
<thead>
<tr>
<th>Year</th>
<th>Contributing institutions</th>
<th>Data categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1953</td>
<td>9 Australian University libraries</td>
<td>Staff, Holdings, Expenditure</td>
</tr>
<tr>
<td>1961</td>
<td>10 Australian University libraries</td>
<td>Staff, Lending, Inter-library loans, Seating, Acquisitions</td>
</tr>
<tr>
<td>1983</td>
<td>26 Australian &amp; New Zealand University libraries &amp; 52 College libraries</td>
<td>Library Staff, Administrative structure, Library Services, Bibliographic Resources, Library Expenditure, Institutional population</td>
</tr>
<tr>
<td>2003</td>
<td>39 Australian Universities 1 Australian archival store 8 New Zealand Universities</td>
<td>Library Organisation, Library Staff, Library Services, Collection Resources, Expenditure, Institutional population</td>
</tr>
</tbody>
</table>

**CAUL STATISTICS FOCUS GROUP**

The CAUL Statistics Focus Group is a committee which oversees the collection and publication of the CAUL Statistics. The Terms of Reference and Operating Guidelines state that ‘the CAUL statistics represent a major achievement of the organisation, and provide a clear, reliable basis for benchmarking, comparison and analysis. They compare favourably with data produced by equivalent bodies in their useability and consistency. They constitute a major resource for the management of university libraries in Australia and New Zealand.’
The goals of the Group are that the CAUL Statistics provide

- Accurate, relevant and authoritative data relating to Australian & New Zealand university libraries
- Consistency of data over time
- Data available widely – accessible, public and widely promoted
- Data which permits analysis of trends
- Collection of both traditional and new information
- Training/feedback opportunities for relevant university library staff
- Statistics which are cheap, useful and valid.

CAVAL works closely with the Group to determine the fields for which information is to be collected and develops and maintains definitions and instructions. It is the role of the Group to ensure that the statistics remain relevant, and capture as much as possible the range of activities and outputs of a university library, so that the development, trial and incorporation of new fields into the collection are regularly discussed. The group also monitors the CAUL statistics ensuring that they are aligned, when appropriate, with other statistical publications, in particular those of ARL, CARL and CURL/SCONUL, as well as international standards, including ISO TC 46 and NISO Z39.7 and the work of the Standards Australia Committee IT/19.

As libraries moved to subscribing to serials in electronic form during the 1990s the collection of statistical data about serials increased in complexity. The completeness and accuracy of the data supplied to CAUL was variable. In 1999, the CAUL statistical definitions were expanded to specifically include several categories of electronic serials and these were further refined in 2000. This has resulted in a discontinuity of data and is a complicating factor when analysing the CAUL serial subscription statistics. An innovative and useful tool, the CAUL Statistics Deemed List, was developed to assist contributors in their annual statistical collections by providing a single source of information for the number of journals in various full text packages, either those from a single publisher, or aggregations of several or hundreds of publishers. The objective is ensure that the same methodology is used by all participating libraries so that appropriate comparisons can be made – rather than being distracted by exactness, for example, whether a particular collection has 543 or 544 titles. The information is collected from publishers, aggregators and web-sites. At this point in time it is acknowledged that it is difficult to de-duplicate titles held in both print and electronic collections.

Incorporation of data on electronic resources into the CAUL Statistics has concerned the Group for several years now. Following the well-publicised ARL E-metrics trial, local trials were run in New Zealand (2003) and Australia (2004). Since 2004 Expenditure on E-resources, has been included in the CAUL statistics data collection. In 2007 a pilot set of questions relating to the reporting of e-books was added.
Contributing libraries are encouraged to supply corrections to data from previous years. Often this occurs when internal processes are reviewed and/or a different staff member takes over responsibility for the statistical data. CAVAL then updates the database and also the full Excel spreadsheet on the CAUL statistics website.

The CAUL Statistics Focus Group encourages institutions to use the CAUL Statistics as their only statistical database. For this reason a number of optional non-core data elements are included. Core data is defined as benchmarkable, required by other bodies or useful for political purposes. It is recommended that non-core (i.e. optional) data is supplied wherever possible, even if only a best estimate, in order to improve the reporting of time series data.

**ASIA ACADEMIC LIBRARIES ONLINE STATISTICS: PILOT PROJECT**

The iGroup (Asia) is providing sponsorship for CAVAL to develop and provide the means for libraries to benchmark regionally across Asia, and eventually internationally. This is being facilitated through the development and implementation of an interactive statistical website for the collection and presentation of statistics for a pilot group of 22 Asian academic libraries.

The participating libraries are

- Hong Kong – Chinese University of Hong Kong, City University of Hong Kong, Hong Kong Baptist University, Hong Kong Institution of Education, Hong Kong Polytechnic University, Hong Kong University of Science and Technology, Lingnan University, The University of Hong Kong
- Malaysia – International Islamic University Malaysia, Universiti Kebangsaan Malaysia, University of Malaya, Universiti Teknologi Malaysia
- Singapore – Nanyang Technological University, National University of Singapore
- Thailand – Chiangmai University, Khon Kaen University, MAEJO University, Mahasarakham University, Silpakorn University, Suan Dusit Rajabhat University, Suranaree University of Technology, Walailak University.

The objectives of the pilot project are to

- develop and provide an online statistical website for Asian academic libraries
- implement sophisticated functionality for online benchmarking
- improve the data collection processes for the individual libraries
- provide an sustainable online statistical service for Asian academic libraries.

Phase 1, during which data for the years 2005 and 2006, has been collected, loaded and made available for benchmarking from the online website is almost complete. Phase 2, the collection of an additional years data (2007), is now underway.
The 3rd phase of the pilot project will be the evaluation of the online website and the service by the participating libraries, the iGroup and CAVAL.

There are a range of benefits to the participating libraries. These include

- local benefits – tracking each individual library over time, developing staff expertise
- institutional benefits – showing the contributions of the library
- national benefits – comparing with other institutions, gaining a national overview of library services
- regional benefits – comparing with similar libraries in other countries, learning from the differences
- global benefits – greater understanding of the role of libraries, opportunities to be involved and contribute to this regional development.

The project began in 2006 with introductory workshops in Singapore, Malaysia, Thailand and Hong Kong. These workshops were generously sponsored by the iGroup (Asia) and explored the use of statistics and other measurement tools to describe, measure, evaluate and benchmark the performance of libraries and their contributions to teaching, learning, research and community service. Following positive feedback to the proposal to setup an Asian Online statistics site based on the CAUL Online Statistics and agreement by the iGroup (Asia) to sponsor the pilot project, initial site development and setup was undertaken.

In 2007 a second set of workshops was run which focussed on the details of participation, including the functionality to be available from the online statistics site and detailed definitions of the data elements for which data is to be collected.

Libraries began entering data into the website in August 2007. Help-desk services for participating libraries are provided by CAVAL (email and telephone).

There are some challenges in developing a statistics service for groups of libraries with cultural and language differences, and physically located across a wide geographic area. Some of the practical issues encountered have been allowing for different currencies and different academic and financial years, and providing appropriate assistance and information for participants with varying technical backgrounds and local infrastructure support. Although the libraries are all keen to be involved and wish to benchmark with CAUL and ARL libraries there is not a tradition of sharing data and hence less familiarity with the practicalities.

FUNCTIONALITY OF THE CAUL AND ASIAN ONLINE STATISTICS SITES

The CAVAL online statistics sites for CAUL and for the Asian Academic Libraries consist of Open Source products (MySQL and Linux). The benefits include no license costs and low hardware specification requirements. However, the cost of developing the ‘front end’ has been higher, as more development time has been
required. CAVAL hosts and manages all aspects of the services – system operation and maintenance including, hardware, operating system, database, web server and network components. CAVAL also provides project management for the development and implementation processes and supplies regular progress reports to participating libraries.

The software developed by CAVAL can be easily adapted to create benchmarking services for any set of statistical data. In particular statistics collected by other cultural institutions such as museums and art galleries and local government agencies.

Both sites provide functionality to

- compare institutional data – up to fifteen institutions can be compared using an unlimited number of variables
- calculate ratios and other statistical measures, displaying results online for all the institutions in the dataset in ranked order.
- display summary statistics – online display of descriptive statistics for all of the institutions in the dataset
- produce graphs – allows the production of an online graph for one institution and up to fifteen variables, or up to fifteen institutions and one variable.
- download data – extract and download a subset of the data by selecting the required institutions, regions (Countries or States), variables, and range of year(s)
- input data – online input of data by staff from contributing institutions, incorporates data validation and online help.

The data for both sites is divided into 6 sections. The 35 variables selected for the Asian Statistics pilot are a subset of the 85 variables currently used for the CAUL Statistics. The sections are

- Library Organisation – general information about the library, including the number of libraries, opening hours, seating facilities, staffing details, shelving and archive capacities
- Library Staff – data about the staff of the library, broken down by both type of staff and position classifications
- Library Services – data about services provided by the library, includes Information Literacy and instruction, Loans, Document Delivery Services, Intercampus and inter-branch lending and turnstile counts
- Information Resources – data about the library’s Information Resources broken down by bibliographic level i.e. serial or non-serial. Non-serial includes monographs and other non-serial works in any medium or format. A pilot set of 4 data elements relating to e-books are included in the collection of the CAUL 2007 data collection
- Library Expenditure – data about the library’s expenditure, broken down into acquisitions, salary and operational expenses
• Institutional Population – includes all staff and students belonging to the institution, including non-academic staff.

CONCLUSION

The CAUL Online Statistics website has become a vital and regularly used tool for identifying and benchmarking data about Australian and New Zealand academic libraries. It is hoped that the Asian Online Statistics site will, over time, expand and develop to provide the same useful service to libraries across the region. The technical design and setup of both sites has been designed to enable future benchmarking between Australian, New Zealand and Asian libraries, and beyond.

CAVAL’s long history of cooperation and collaboration with libraries, combined with more recently developed technical and statistical expertise, provides a solid platform from which to further develop online statistical and benchmarking services. There are exciting opportunities for international collaboration which will be welcomed by many libraries.

REFERENCES

CONZUL (Council of New Zealand University Librarians) – http://www.conzul.ac.nz
iGroup – http://www.igroupnet.com

SCONUL (Society of College, National and University Libraries) statistics – http://www.sconul.ac.uk/statistics/